

ASSESSING STUDENT ACHIEVEMENT:
MULTIPLE-CHOICE TESTING VERSUS PERFORMANCE ASSESSMENT

BY

JANA C. DEMING

AN APPLIED RESEARCH PROJECT
(POLITICAL SCIENCE 5397)
SUBMITTED TO
THE DEPARTMENT OF POLITICAL SCIENCE
SOUTHWEST TEXAS STATE UNIVERSITY
IN PARTIAL FULFILLMENT
FOR THE REQUIREMENTS FOR THE DEGREE OF
MASTERS OF PUBLIC ADMINISTRATION

(FALL 1992)

FACULTY APPROVAL:

Pat Shields

Carol Kew

TABLE OF CONTENTS

	Page
List of Tables	iv
CHAPTER ONE. INTRODUCTION	
The Rise of Standardized Testing	2
Multiple-Choice Testing and Performance Assessment ..	4
Purpose of the Research	6
Chapter Summaries	6
CHAPTER TWO. REVIEW OF LITERATURE	
Introduction	8
Test Design	8
Impact on Teaching and Curriculum	13
Impact on Students	17
Equity in Testing	21
Cost	27
Time	31
Conclusion	34
CHAPTER THREE. TESTING IN TEXAS	
Historical Overview	36
High Stakes Assessment	38
Public Reaction to Statewide Assessment	40
Student Performance	42
The Future of Testing in Texas	43
CHAPTER FOUR. METHODOLOGY	
Introduction	46
Methodology	46
Sample Selection	47
Survey Design	48

Categories of Issues	49
Issues/Survey Questions	53
Statistical Methods	54
Strengths and Weaknesses of Survey Research	55
CHAPTER FIVE. RESULTS	
Test Design	58
Impact on Teaching and Curriculum	60
Impact on Students	63
Equity in Testing	65
Cost	67
Time	68
Demographic Information	69
CHAPTER SIX. CONCLUSION	
Conclusion	73
APPENDIX	
A. Summary Tables 1-6	78
B. Survey Instrument	81
C. Letter to District Test Coordinators	88
BIBLIOGRAPHY	89

TABLES

Table	Page
4	Issues/Survey Questions54
5.1	Test Design58
5.2	Test Design59
5.3	Impact on Teaching and Curriculum60
5.4	Impact on Teaching and Curriculum62
5.5	Impact on Students63
5.6	Equity in Testing65
5.7	Cost67
5.8	Time68
5.9a	Demographic Information69
5.9b	Demographic Information70
5.10a	Demographic Information70
5.10b	Demographic Information70

CHAPTER ONE

INTRODUCTION

Student assessment in American public schools is today a topic of considerable debate. Standardized test results affect not only the school careers of children, but the careers of teachers and education administrators as well. The impact of standardized test results is felt at local, state, national, and international levels.

Ruth Mitchell defines a standardized test as one which is "given under the same conditions and asks the same questions across different populations in order to permit comparisons" (Mitchell, 1992: 5). Mitchell describes assessment as a tool for placing students in instructional programs, improving the teaching and learning process, and achieving school accountability (Mitchell, 1992: 19-22). According to Archbald and Newmann, standardized testing is also useful in selecting students for college and financial assistance, and in recognizing students, teachers, and schools (Archbald & Newmann, 1988: v).

The use of standardized testing today is extensive. According to the National Commission on Testing and Public Policy, 41 million children in U.S. public schools take an estimated 127 million tests annually. Furthermore, this estimate is probably somewhat conservative considering that students in special programs, such as those financed by

federal funds, are administered additional tests in order to meet federal reporting requirements (Mitchell, 1992: 4).

THE RISE OF STANDARDIZED TESTING

The importance attached to standardized testing has increased dramatically over the last 20 to 30 years. This is the result of a series of changes that evolved over many years. Since the 1960s, social change in the U.S., such as the civil rights movement and the women's movement, has produced a call for quality and equality of opportunity for all citizens (Airasian, 1987: 395-401).

The impact of social change on public education can be observed in increases in school control and school growth, and in the increased politicization of education. Concern regarding the quality of education has increased school control, and caused a shift in focus from the local level to the state and national levels. The press for equality has required that schools respond to students with special needs such as limited English proficient or dyslexic students. Schools have grown as they have been required to implement new programs to fulfill additional goals and functions. The focus on equality has also caused public education to become more political because of the need to satisfy powerful special interest groups, and resolve controversies regarding the distribution of limited resources (Airasian, 1987: 395-401).

Education became a national issue in the 1970s as public doubt mounted regarding the effectiveness of American schools. This doubt was heightened by the publication of *A Nation At Risk: The Imperative for Education Reform*, a 1983 report by the National Commission on Excellence in Education. The report labelled the American educational system inadequate, and charged that it had failed to keep pace with world change following World War II. According to the Commission, schools were failing to teach students the skills required to compete in a global economy. Their report called for the teaching of higher-order thinking skills, problem solving, cooperative work theory, and the development of ideas (Mitchell, 1992: 174-175).

The 1980s brought additional demands for education reform. Critics of the American educational system called for "restructuring American schools, professionalizing teachers, reshaping curriculum toward concepts and application, and the reform of testing" (Mitchell, 1992: 175).

In addition, the American public has generally been supportive of standardized testing. To many, testing represents traditional educational standards and values. Tests are considered fair and objective instruments whose scientific value is verified through a numerical score. Furthermore, testing provides legislative control over schools. Parents, legislators, and even the judiciary seem

to prefer standardized tests to teacher assessment as a measure of student achievement (Airasian, 1987: 394-407).

It is thus evident that a series of changes has created an educational system driven in large part by standardized testing. Standardized test results are a major criteria in the school decision making process. They are used to select students for school programs and determine student placement. In addition, standardized test results are used to measure the performance of teachers, administrators, schools, and school systems. Ultimately, standardized test results have become the yardstick by which success or failure is measured (Neill & Medina, 1989: 688).

MULTIPLE-CHOICE TESTING AND PERFORMANCE ASSESSMENT

Although a variety of measures exist for assessing student achievement, multiple-choice testing has been the most commonly used method for the last 40 years (Feinberg, 1990: 14). Test questions are generally short, machine scorable items which require students to select answers from among several predetermined answer choices (Mitchell, 1992: 172).

Multiple-choice testing has come under fire in recent years. Critics charge that multiple-choice tests ignore important areas of school curriculum, and focus instead on basic skills, rote memorization, and short-term recall (The University of California at Los Angeles, 1990: 4-5).

The significance of conducting quality assessment is that testing drives instruction to a large degree. The pressure to raise standardized test scores pushes teachers to "teach to the test"; in other words, what is tested is what is taught. Unfortunately, teaching methods that are effective in raising multiple-choice test scores are often boring and uninspiring, and deemphasize important areas such as higher-order thinking skills and cooperative learning. Accordingly, current education reform stresses the importance of improved assessment in the improvement of education (The University of California at Los Angeles, 1990: 4-5).

The press for improved student assessment has spawned national interest in performance assessment, a method of testing praised as a measure of true learning. The most obvious difference between performance assessment and multiple-choice testing is that performance assessment requires students to construct original responses which are scored by hand. Also referred to as authentic or direct assessment, performance assessment is defined as "an exercise in which examinees apply the skills and knowledge they have mastered by demonstrating specific skills and competencies" (Stiggins, 1987: 34). Examples of performance assessment include drawing or painting, conducting science experiments, measuring or using computers in mathematics, using reference materials in reading, and writing essays and letters (Roeber, 1989: 4).

PURPOSE OF THE RESEARCH

The purpose of this research is twofold. First, the research will compare two methods of standardized testing: multiple-choice testing and performance assessment. By examining the expert opinion of educators across the nation, the issues regarding each method of student assessment will be defined.

Second, the attitudes of Texas educators toward the two methods of assessment will be examined. The educators targeted in the study are uniquely qualified to address the issues regarding student assessment. As school district test coordinators, these educators are experienced in both teaching and standardized testing.

CHAPTER SUMMARIES

Chapter One defined standardized testing, multiple-choice testing, and performance assessment. The process by which standardized testing has become an important component in the American educational system was described. Finally, the purpose of the research was explained.

Chapter Two will present a review of the literature on multiple-choice testing and performance assessment. Because little, if any, empirical evidence exists which compares multiple-choice testing to performance assessment, the literature review will focus on expert opinion. The major issues involved in student assessment will be addressed.

These issues are categorized as follows: test design; impact on teaching and curriculum; impact on students; equity; cost; and time. The underlying facets within each category will be discussed.

Chapter Three will provide an historical overview of the statewide testing program for Texas. The high stakes attached to student assessment in Texas and their impact on students, teachers, and schools will be addressed. This chapter will also describe the state's shift toward an assessment program that is primarily performance based.

Chapter Four will discuss the characteristics of survey research. The sample selection, survey design, and pre-test will be detailed. Statistical methods used to measure the survey data will be described. The strengths and weaknesses of survey research will be discussed.

In Chapter Five, the survey results will be analyzed. The results will reflect in detail the responses of Texas educators regarding multiple-choice testing and performance assessment. The results will highlight differences in attitudes regarding the two methods of testing.

Chapter Six will conclude the Applied Research Project by summarizing the attitudes of Texas educators toward multiple-choice testing and performance assessment. The trade-offs that emerge in the use of each testing method will be discussed.

CHAPTER TWO

INTRODUCTION

The literature which compares and assesses multiple-choice testing and performance assessment consists of the expert opinion of American educators. There is little, if any, empirical evidence to support or refute expert opinion. Nevertheless, the arguments of these experienced educators are persuasive.

This research is descriptive in nature. The approach to the research question will combine a review of the current literature on student assessment with a survey of education professionals.

The issues involved in student assessment may be classified within six broad categories. These categories are test design, impact on teaching and curriculum, impact on students, equity in testing, cost, and time. Each category contains within it a number of facets which will be discussed as they relate to multiple-choice testing and performance assessment. Tables which display the categories and their underlying facets appear in Appendix A.

Test Design

This section will address the facets of test design that are of concern to educators. These facets include

skills measured, range and focus of coverage, flexibility for demonstrating achievement, and scoring.

Skills measured

In Lawrence Feinberg's opinion, it is important that students acquire a solid foundation of basic skills on which to build. He claims that multiple-choice tests of basic skills measure both content knowledge and academic ability required for college. He attributes the recent improvement in basic skills test scores to the increase in basic skills testing, and reminds educators of the importance of this accomplishment (Feinberg, 1990: 14-17). Grant Wiggins argues, however, that multiple-choice tests of basic skills are unchallenging and unrealistic in light of the skills required of students in the real world (Wiggins, 1991: 702). These tests are considered by many to be poor predictors of student ability (Haney & Madaus, 1989: 684).

In contrast, performance assessment is praised for its emphasis on higher-order thinking skills, problem solving, and the measurement of process and effort. It is touted as realistic, meaningful, and instructional (The University of California at Los Angeles, 1990: 5-6). Doug Archbald and Fred Newmann applaud performance assessment for its use of "disciplined inquiry, integration of knowledge and value beyond evaluation" (Archbald & Newmann, 1988: 2).

Performance assessment is not without its critics, however. Some experts doubt that it measures anything

different than multiple-choice tests. To the claim that performance assessment measures higher-order thinking skills, and is "nonalgorithmic, complex, yields multiple solutions and nuanced judgement", expert response is often skeptical (Cizek, 1991: 698). Critics agree that it sounds good, but insist that what will really be measured and how is not sufficiently explained (Cizek, 1991: 698).

Range and focus of coverage

Multiple-choice tests offer a broader range of topic coverage than performance assessments. Because time constraints often force teachers to eliminate areas not tested, many experts in student assessment feel that a broad range of topic coverage is critically important (Linn, Baker, & Dunbar, 1991: 20). These experts maintain that the narrower range of performance assessment may unfairly emphasize topics with which students are unfamiliar (Feinberg, 1990: 17), and may not be truly representative of students' knowledge and ability (Arter, 1991: 4).

Although multiple-choice tests offer a broader range of topic coverage, test publishers are often criticized for limiting the depth of multiple-choice tests by matching them to a small number of popularly used textbooks. The focus on basic skills is perpetuated, and reading passages in tests are shorter and easier than those found in textbooks. To avoid excessive changes in test format, only a few types of items are used despite the fact that students might be able

to respond correctly if the items were presented in another manner (Shepard 1989: 5).

Despite its narrower range of topic coverage, performance assessment is touted for its in-depth focus. For example, performance assessments often encompass more than one performance. Students' work is assessed repeatedly over time utilizing techniques such as portfolios which involve many demonstrations of achievement. It is thus possible to observe patterns and reasons for success or failure (Wiggins, 1989: 705).

Flexibility for demonstrating achievement

Critics of multiple-choice testing claim that multiple-choice achievement tests measure isolated and superficial fragments of knowledge without offering students an opportunity to demonstrate the ability to integrate knowledge (Archbald & Newmann, 1988: 60). Furthermore, these tests are considered to be unrealistic because students have no opportunity to defend or explain wrong answers (Wiggins, 1989: 712).

Performance assessment recognizes that students may have a justifiable need to have questions phrased in another manner, and that students can often demonstrate ability with only minimal assistance. It affords students an opportunity to explain and discuss their answers, and to obtain feedback from test administrators. Tests may allow for different levels of difficulty so that all students, regardless of

skill or experience, can demonstrate their achievement (Wiggins, 1989: 712).

Supporters of performance assessment hold that the human interaction it offers is realistic and invaluable. In contrast to the inflexibility of machine scored tests, performance assessment offers the interaction between students and test administrators that is required in order for students to be fairly and equitably judged (Wiggins, 1989: 708).

Scoring

One of the major advantages of multiple-choice tests is that they are considered fair and objective because they are machine scorable. Because performance assessments are scored by hand, many experts fear the potential for inconsistency in scoring (Feinberg, 1990: 17) and the possibility that results will be affected by the bias of individual scorers. It is feared that teacher scoring will result in unacceptable levels of subjectivity (Frechtling, 1991: 25). In addition, newly developed performance assessments may not include adequate criteria for scoring (Arter, 1991: 4).

On the other hand, Wiggins claims that the design of multiple-choice tests can still allow room for subjectivity. In his opinion, performance assessment can achieve acceptable levels of objectivity if scorers are trained in the use of uniform scoring criteria. Furthermore, the use

of multiple judges can achieve high levels of inter-rater reliability (Wiggins, 1989: 710).

Impact on Teaching and Curriculum

Educators are under a great deal of pressure to produce students who perform well on standardized tests. As testing drives instruction, a number of issues emerge regarding teaching and curriculum. This section will discuss skills taught, teaching methods, curriculum content coverage, and textbook content coverage.

Skills taught

Critics of multiple-choice testing claim that undue reliance on basic skills tests has corrupted teaching by forcing teachers to focus on teaching basic skills to the exclusion of other valuable skills and abilities (Haney & Madaus, 1989: 684). In an effort to improve test scores, teachers focus on the memorization and recall of isolated facts. Thus, students are not required to demonstrate true understanding (Mitchell, 1992: 15).

In contrast, performance assessment is thought to have a positive impact on the skills taught to students. For example, surveys of educators in English and language arts have produced overwhelming evidence that multiple-choice tests discourage writing instruction while the use of writing samples encourages it. Although writing samples are more time-consuming, they have a tangible and favorable

impact by encouraging the teaching of writing skills (Haney & Madaus, 1989: 686).

Yet there are those who claim that improving basic skills is important in and of itself, and that teaching to the test has merit when one considers the improvement in basic skills achievement of the last decade. Some experts claim that properly designed and interpreted multiple-choice tests can adequately predict whether students have acquired the skills necessary for successful performance on the job and in college. Finally, no performance assessment will prove meaningful to a student who doesn't have at least some basic knowledge (Feinberg, 1990: 14-17).

Teaching methods

In general, test publishers design tests that assess skills which are easily measured. Teaching follows the same pattern to the exclusion of skills that are more difficult to assess (Mitchell, 1992: 15). Teachers employ lectures and worksheets in the hope of improving students' scores on multiple-choice tests of basic skills (Darling-Hammond, 1991: 222). Unfortunately, teaching methods appropriate for teaching basic skills bear little resemblance to those required for the teaching of higher-order thinking skills (Neill & Medina, 1989: 693-695).

Probably the strongest claim for the superiority of performance assessment is that it has a positive impact on teaching methods and curriculum by forcing the teaching of

the kinds of skills our students need most (Wiggins, 1991: 703). Performance assessment encourages the use of teaching methods that include writing, research, student discussion, science experiments, and cooperative learning activities (Darling-Hammond, 1991: 222).

Critics claim, however, that performance assessment may be subject to the same ills as multiple-choice testing. For example, in Maryland, schools are told in advance the structure to be used on the essay required for high school graduation. Teachers then teach the essay formula, and most students perform well on the first attempt. Unfortunately, the impact on writing outside the formula is minimal (Feinberg, 1990: 17).

Curriculum content coverage

One of the primary disadvantages of standardized multiple-choice tests is that they do not reflect school curriculum. These tests are usually national norm-referenced tests which often assess students in areas unfamiliar to them while failing to test students in areas with which they are familiar (Archbald & Newmann, 1988: 58-59). School curriculum is therefore determined to a significant degree by test publishers rather than by local school districts (Airasian, 1987: 406). In addition, curriculum is often narrowed to reflect the teaching methods used to bolster student test scores (Neill & Medina, 1989: 693-695).

In contrast, performance assessments are thought to have a positive impact on curriculum by broadening the range of skills emphasized and by fostering an emphasis on higher-order thinking skills. For example, California's "first-class assessment" and New York's hands-on science experiment have led to an increased focus on writing and experimentation in their public school systems (Wiggins, 1991: 703).

There are those, however, who question the assertion that multiple-choice tests are inherently simplistic and performance assessments inherently challenging. According to Feinberg, the decision to challenge students is a political and educational decision made by school districts, and is unrelated to test format. In his opinion, multiple-choice tests have the potential to challenge students. For example, the National Assessment of Educational Progress (NAEP) tests a broad range of curriculum and promotes broad learning. By the same token, performance assessments may be simplistic (Feinberg, 1990: 16-17).

Textbook content coverage

Critics of multiple-choice tests claim that the tests have caused not only a narrowing of school curriculum, but have resulted in the narrowing of textbooks as well. With the introduction of standardized tests in the 1920s, good literature began to disappear from textbooks because standardized tests emphasized reading skills rather than

reading itself. Some experts in education contend that the importance attached to standardized tests coupled with attempts by schools to align their curriculum with the tests are responsible for a decline in the quality of textbooks. As a result of watered down curriculum and textbooks, even the best students may lose interest and perform poorly (Neill & Medina, 1989: 693-695). In contrast, supporters of performance assessment feel that it will have a positive impact on curriculum and textbooks by broadening the range of skills emphasized (Wiggins, 1991: 703).

Impact on Students

The use of standardized test results has a tremendous impact on students. This section will address issues surrounding the use of test results for program placement and promotion/retention. It should be noted that most experts in education agree that standardized test results should not constitute the sole basis for making decisions regarding students. This section will also address whether testing is meaningful for students.

Program placement

The placement of students in instructional programs, commonly referred to as "tracking", based solely on multiple-choice basic skills test results has come under fire in recent years. Critics of this practice charge that tracking students using test results allows for an increased

risk of misdiagnosis (Darling-Hammond, 1991: 222-223), and is most harmful to students in the lower tracks (Neill & Medina, 1989: 693). The standards for these students never converge with those of the higher tracks. Students in the lower tracks do not receive the instruction or assessment necessary to help narrow the gap between them and their peers. It is simply easier for students in the lower tracks to earn better grades (Wiggins, 1989: 711).

Some experts maintain that the information provided by performance assessment is "far richer, more complex, and more powerful" than that provided on traditional tests (Neill & Medina, 1989: 696). For diagnostic purposes, essays and free response items are much more informative than multiple-choice tests. Performance based assessments reveal how answers are derived and how errors are made (Feinberg, 1990: 31).

According to some experts, however, multiple-choice testing is unfairly maligned. Lawrence Feinberg argues that it makes little difference what type of test is used for the purpose of program placement. For example, on the California Bar exam, individuals who score poorly do so on all parts of the exam; on the performance assessment, the essay, and the multiple-choice portion. Those who score well on one portion score well on all portions. The same is true of the Advanced Placement computer science exam. Examinees perform uniformly on the free response portion and

the multiple-choice portion of the exam (Feinberg, 1990: 31).

Promotion/retention

Test results are often used to determine graduation eligibility (Airasian, 1987: 404). The issue of misdiagnosis is important in making both program placement and promotion/retention decisions. Although it is commonly accepted that some students test poorly for reasons other than lack of ability, students are often placed in remedial programs based on poor performance on standardized multiple-choice tests. The focus of these programs on basic skills leaves many students bored and frustrated. As students fail to improve, the focus on basic skills increases, and the likelihood of students dropping out also increases (Neill & Medina, 1989: 693-695).

Multiple-choice tests have been criticized for providing results that are often inaccurate and inconsistent measures of achievement, ability, and skills (Neill & Medina, 1989: 689). The tests fail to produce accurate information about the state of achievement in our schools (Haney & Madaus, 1989: 684).

In contrast, it is claimed that performance assessment provides accurate information about what students know, and that it effectively measures student progress. These assertions are based on the fact that authentic assessment requires students to actively demonstrate or perform. Ruth

Mitchell contends that if we want students to be able to "analyze, interpret, synthesize, and evaluate, progress should be charted through a direct performance" of these skills (Mitchell, 1992: 20-21).

Supporters of multiple-choice testing, however, cite a 1982 National Academy of Sciences committee which stated that these tests are good predictors of both job and college performance if properly developed and interpreted. Especially for college, it is important to know where students stand academically, and the tests clearly fulfill this function. Students require basic knowledge in order to reason about a subject, and tests of basic skills are a good method for determining what students know (Feinberg, 1990: 14-16).

Meaningful assessment

Multiple-choice tests are often criticized for measuring learning which is trivial and meaningless (Archbald & Newmann, 1988: 1). Critics claim that multiple-choice testing measures skills out of context, and leads to endless drill and practice (Shepard, 1989: 5). It has been suggested that a link exists between low performance and the extent to which students find tests to be meaningful (Linn et al., 1991: 20).

Performance assessment is thought by many experts to offer a meaningful educational experience in a contextualized setting (Linn et al., 1991: 20). Current

learning theory holds that students learn best when they are able to make the link between facts and concepts (Shepard, 1989: 5-6). Opponents of multiple-choice testing contend that good assessment has value beyond evaluation, and allows students to produce a product and cooperate with others in a flexible time frame (Archbald & Newmann, 1988: 3).

According to the U.S. Department of Labor's 1991 Secretary's Commission on Achieving Necessary Skills (SCANS) report, "What Work Requires of Schools", assessment should form a bridge between school and the real world. One suggestion is to create nationally accepted certificates of achievement based on performance tasks, portfolios, and projects. It is thought that certificates recognized nationwide by employers would motivate students by linking "learning and earning" (Packer, 1992: 30).

Equity in Testing

This section will address equity in testing. Factors discussed will include equity in assessing minority students, low-income students, and limited English proficient students. The allowance of testing for varied learning styles will also be discussed.

In all fairness, it should be noted that the equity of performance assessment is as yet largely undetermined. What follows is a summary of expert opinion which supports the equity of performance assessment. Afterward, discussion of the equity of multiple-choice testing is presented.

Proponents of performance assessment maintain that alternative forms of assessment, such as performance-based portfolios, show what students know as well as the different ways in which they learn (Neill & Medina, 1989: 695-696). In addition, Wiggins argues that human judgement should not be replaced with machine scoring. In real life, people are allowed to change their responses after having a question rephrased. Furthermore, human judges are utilized in athletic events and legal proceedings because it is not possible to "reduce complex judgements to rules" and still maintain equity (Wiggins, 1989: 708). Even on well developed multiple-choice tests, students may need to have questions rephrased and have a chance to explain their answers (Wiggins, 1989: 708).

Still, education experts warn that care must be taken not to replace the biases of current tests with those of teachers and schools. If, as expected, the information yielded by performance assessments is better, we must be careful in the use of results. Minorities and low-income students will continue to be most vulnerable to the misuse of test results (Neill & Medina, 1989: 696).

Minority students

One of the most significant criticisms of standardized multiple-choice tests is that they fail to equitably assess the achievement of minority, low-income, and limited English proficient (LEP) students. These tests reflect the

language, culture and lifestyle of white middle and upper class students, and the topics assessed are often those with which some minority students are unfamiliar. The tests assume that students share the same experiences. Critics charge that standardized tests are really a measure of race and income rather than achievement (Neill & Medina, 1989: 690-692).

Some experts claim that much of the disparity in the achievement of white students and minority students is due to the use of test results for program tracking. Minority students are disproportionately more likely to score poorly on standardized tests, and therefore are often placed in remedial programs that offer unchallenging curriculum. As a result, their achievement continues to remain low (Darling-Hammond, 1991: 222).

Other characteristics of multiple-choice tests hurt minority students as well. Nonstandard test administrations and unfamiliarity with test administrators tend to have a negative impact on the performance of minority students. Unlike white middle class students, minority students are likely to achieve lower scores when they do not know test administrators. Furthermore, test time restrictions hurt the performance of black and Hispanic students (Neill & Medina, 1989: 690-692).

Some critics of multiple-choice testing contend that performance assessment offers a more equitable measure of the achievement of minority students. There is some

evidence, however, that performance assessment is not as equitable toward minority students as its supporters maintain. While there is little research which compares test scores by race on multiple-choice tests versus performance assessments, the 1984 addition of a written performance assessment to the California Bar exam produced interesting results. Although the pass rate for all groups fell, the gap between blacks and whites did not narrow. In fact, a study by the Rand Corporation predicted that adjusting the scores to reflect the lower reliability of performance assessment would actually show a wider gap between the two groups than evidenced on the multiple-choice test (Feinberg, 1990: 15-16).

Another point of interest is the effect of time allowances on multiple-choice tests as compared to performance assessments. In its 1988 writing exam, the National Assessment of Educational Progress (NAEP) showed that the gap in performance between races widened when the time allowed for each question was increased from 15 to 30 minutes. It is important to note that the racial gap had been about the same on shorter writing tests as on the multiple-choice portions of the NAEP exams. When the time allowance was increased, the average performance of whites substantially improved while the performance of blacks and Hispanics was little changed. According to the NAEP, the extra time allowed good writers more time to organize and polish their writing, but it did not help those whose

writing ability was limited at the outset. Because disproportionately more of the good writers were white, more whites than minorities showed improved performance. South Carolina witnessed the same results a few years earlier when the time allowed for each writing question was increased (Feinberg, 1990: 15-16).

Low-income students

Standardized tests negatively impact low-income students in many of the same ways that they impact minority students. The tests assume middle and upper class cultural experiences which may not be shared by students from low-income families. The performance of low-income students also suffers when nonstandard test administrations are used and when students are unfamiliar with test administrators (Neill & Medina, 1989: 690-692).

Again, some critics of multiple-choice tests contend that the use of test results for program tracking explains much of the disparity in achievement between middle and upper class students and low-income students (Darling-Hammond, 1991: 222).

Limited English proficient students

Standardized tests pose special problems for students with limited English proficiency (LEP). Bilingual students switch repeatedly from one language to another in both speech and thought patterns. Because LEP students require

more time to process information in a second language, their test scores are greatly impacted by timed standardized tests. LEP students' test scores also suffer because the use of a non-native language causes them to be more easily distracted by noise than are native English speakers. In addition, students who hear a language other than English in their homes are likely to receive test scores which inaccurately assess their academic achievement. For these reasons, many experts in bilingual education contend that tests in the English language measure English proficiency rather than subject knowledge (Ascher, 1991: 7-8).

Perhaps the most serious problem in the standardized testing of LEP students is that low scores are often considered to be indicators of learning deficits or disorders. There is evidence that a disproportionate number of Hispanic students are labelled as mentally retarded based solely on the results of multiple-choice intelligence tests (Ascher, 1991: 7). Furthermore, the importance attached to test scores provides an incentive to place low scorers in special education programs because special education students' scores are not aggregated with the schools' overall results. This practice is most damaging to LEP students and to minority and low-income students as well (Darling-Hammond, 1991: 223).

Allowance for varied learning styles

Although recent research in child development indicates that children develop in different ways and at varying paces, traditional assessment fails to account for normal variations. Multiple-choice tests are criticized because they assume that all students share identical styles of learning and problem solving (Neill & Medina, 1989: 689-692).

Authentic performance based assessment is thought to accommodate varied learning styles, interests, and aptitudes. According to Wiggins, performance assessment closely resembles real life by allowing for variation in the approach to tasks, topics, and procedures. In addition, slower response times may be accommodated (Wiggins, 1989: 712).

The interest in performance assessment has been spurred not only by progress in the knowledge of cognitive science and the learning process. Improved computer and information technology has also added fuel to the fire. For example, computer adaptive testing has been developed which has the ability to match the difficulty level of individual students (Haney & Madaus, 1989: 685).

Cost

The costs associated with testing extend far beyond the cost of printing test booklets and answer documents. This

section will address the costs associated with test development, equipment/materials, and scoring.

Development

No reliable method currently exists for comparing the cost of multiple-choice testing to performance assessment. However, it is estimated that the overall cost of multiple-choice testing is approximately \$1.50 per examinee as compared to \$3-10 per examinee for performance assessment (The University of California at Los Angeles, 1990: 6).

One of the primary advantages of multiple-choice testing is the relatively low cost of test development. The boom in the production of these tests during the 1960s and 1970s provided a wealth of multiple-choice tests (Mitchell, 1992: 172). Performance assessments, however, are not abundant, and are more costly to develop because their development is more complex and time-consuming. Critics of performance assessment fear that if funds for testing are drawn from other areas, such as staff development, infighting over the distribution of resources will result (Cizek, 1991: 697).

Nevertheless, it is possible to minimize the cost of development for performance assessment. For example, in Michigan's Physical Fitness tests, the state education agency staff was supplemented by a core group of volunteers composed of teachers, curriculum specialists, college educators, and volunteers from subject area organizations.

Twenty tests and scoring guides were developed for a mere \$10,000 (Roeber, 1989: 6).

Equipment/materials

Performance assessment is more costly than multiple-choice testing because it may require additional equipment and materials. While multiple-choice tests require schools to supply only No. 2 pencils, performance assessment may require the use of libraries and laboratory equipment (The University of California at Los Angeles, 1990: 10-16).

The cost of performance assessment, however, can be reduced. Samples of students may be tested instead of testing all students. Also, some states are testing students in pairs or groups, and others assess different content areas in different years. It should be noted, however, that some state legislatures require all students to be tested in specified content areas. For those states, the use of sampling and rotating content areas are not possible. In addition, diagnostic information may be required for certain groups of students. For example, federal funds for Chapter 1 and Title I programs are allocated to districts and schools for students who fall below state or locally established cutoff points. Students must be administered an approved nationally aggregable multiple-choice test each year to measure their progress and to determine whether funding will continue (Mitchell, 1991: 7-8). Thus, schools with large numbers of Chapter 1 or

Title I students may be unable to fund one testing program for students with special needs and another for the general population (The University of California at Los Angeles, 1990: 10-11).

Scoring

It is estimated that the overall cost of performance assessment is 2-7 times greater than that of multiple-choice testing (The University of California at Los Angeles, 1990: 6). The biggest cost advantage of multiple-choice testing is in the area of scoring. Multiple-choice testing gained in popularity during the 1960s and 1970s largely because of advances in computer technology which allowed for inexpensive machine scoring of test results (Mitchell, 1992: 172).

Scoring is one of the most costly aspects of performance assessment because each test must generally be scored by hand. Some states employ professional scoring services while others use teachers as scorers. Although teacher scoring provides the benefit of professional development for teachers, some experts feel that the expense of teacher scoring is too great. Funding must allow for the cost of training teachers to score, travel reimbursement to training sites, paying teachers for scoring results, and hiring substitute teachers to replace those engaged in scoring (The University of California at Los Angeles, 1990: 8-9).

Nonetheless, some states developing performance assessments are working on methods for reducing the cost of scoring. Some have determined that professional scoring services offer the most cost-efficient alternative to teacher scoring. Others are experimenting with remote scoring via computer networks, collaborative efforts with local education agencies, and the use of college volunteers (The University of California at Los Angeles, 1990: 8-9).

Time

This section will address time-related issues in standardized testing. These include test development, test administration, and scoring.

Test development

The time required for test development is not an issue in multiple-choice testing. As scantrons and computers became cheaper and easier to use in the 1960s and 1970s, technology revolutionized standardized testing. The test publishing industry experienced massive growth, and the multiple-choice norm-referenced test became the test of choice. After a long history of use, there is an abundance of widely accepted tests (Mitchell, 1992: 172).

The development of new assessment strategies always requires more time than the use of established measures. In addition, the nature of performance assessments makes them more time-consuming to develop. Development is more labor

intensive because of the need to train state staff on new strategies, train test administrators, train scorers, and train and orient educators (The University of California at Los Angeles, 1990: 7). Nevertheless, development time can be decreased by using volunteers, such as teachers, professional educator groups, college professors and graduate students, to assist state agency staff (Roeber, 1989: 6).

Test administration

One of the primary concerns of educators regarding testing is the amount of instructional time lost to testing. Proponents of multiple-choice testing emphasize the ease and speed with which tests can be administered (Archbald & Newmann, 1988: 52). In contrast, performance assessments are more time-consuming to administer. Critics complain that some performance assessments require excessive paper routing and additional materials, and are therefore too complicated (The University of California at Los Angeles, 1990: 10-16).

One time-saving alternative in performance assessment is the use of sampling instead of testing all students. Other options include testing students in group settings, alternating the content areas tested from year to year, and the use of integrated assessments which test more than one subject area at a time. Unfortunately, sampling is not possible if state legislatures require that all students be

tested (The University of California at Los Angeles, 1990: 10-11). In addition, diagnostic multiple-choice tests may be required to satisfy reporting requirements for students in federally funded programs (Mitchell, 1992: 7-8).

Scoring

A primary advantage of multiple-choice testing is the efficiency of scoring. Because the tests are machine scorable, scoring is quickly and easily accomplished (Archbald & Newmann, 1988: 52).

In comparison, performance assessments require time-consuming scoring by hand, and if teachers are employed as scorers, they must spend valuable time away from their classes. The time required for the analysis and reporting of results is also increased. Therefore, the use of performance assessment with large student populations may make it difficult for schools to obtain assessment data in a timely fashion. As a result, schools may instead opt for simplified scoring and analysis (The University of California at Los Angeles, 1990: 8-11).

On the other hand, the time required for scoring performance assessments can be decreased by using professional scoring services (The University of California at Los Angeles, 1990: 8). Another method, demonstrated in the Michigan Physical Fitness tests, is to supplement teacher scoring with volunteers. In the Michigan tests, the

scoring of 20 tests for 1600 students was completed in one or more two-day scoring sessions (Roeber, 1989: 7).

CONCLUSION

It is evident from the literature review that student assessment is an important part of the educational process. It is also apparent that multiple-choice testing and performance assessment are markedly different, each offering advantages and disadvantages.

It is difficult to weigh the relative merit of the arguments offered by education experts. One's perspective is certainly influenced by one's position. For example, an employer seeking qualified employees would presumably be quite concerned with the types of skills acquired by students, while a bilingual teacher is likely to be more concerned with the fairness of standardized tests to limited English proficient students. To understand the issues involved in student assessment, it may be more effective to summarize the most persuasive arguments regarding each method of testing. The major arguments which have been developed in this chapter supporting and refuting each testing method appear in Summary Tables 1-6 in Appendix A.

The review of the literature on multiple-choice testing and performance assessment has failed to yield identifiable hypotheses. This may be in part because education experts offer numerous convincing arguments and counterarguments for many of the issues related to standardized testing.

Furthermore, it would be difficult to form hypotheses regarding the attitudes of Texas educators based on the opinions of educators from around the nation. It is hoped instead that the survey results will generate hypotheses, and perhaps serve as a foundation for future research.

Before discussing the survey and its results, it is appropriate to examine student assessment in Texas. The following chapter will address the statewide student assessment program and its impact on education in Texas.

CHAPTER THREE

HISTORICAL OVERVIEW OF TESTING IN TEXAS

State mandated student assessment began in Texas in 1980 with the administration of the Texas Assessment of Basic Skills (TABS). Students in Grades 3, 5, and 9 were tested in reading, writing, and mathematics. The TABS was a criterion-referenced test (CRT), or one that is designed to measure individual performance on a particular set of criteria. Its purpose was to measure student mastery of the Texas school curriculum (Texas Education Agency, 1990: Foreward).

In 1985, the TABS was replaced by the Texas Assessment of Minimum Skills (TEAMS). Also a CRT, the TEAMS assessed reading, writing, and mathematics in Grades 1, 3, 5, 7, 9, and 11. (Testing at Grade 1 was later discontinued by the Texas Legislature.) For the first time, students in Texas public schools were required to pass a state mandated basic skills test in order to be eligible to receive a high school diploma. High school diplomas were first withheld based on statewide test results in May 1987 (Texas Education Agency, 1990: Foreward).

With each new testing program, the tests were expanded to encompass additional instructional objectives. Although the difficulty of the tests was increased, student performance steadily improved. The testing program had not only measured student achievement, but had raised student

achievement levels across the state (Texas Education Agency, 1990: Foreward).

Nevertheless, it soon became evident that the TEAMS minimum skills had become the maximum skills. School curriculum narrowed as the test became the primary focus of instruction. Teachers, curriculum supervisors, and school administrators from around the state attributed the emphasis on basic skills to the growing uses of TEAMS results and the pressure to raise test scores (Texas Education Agency, 1990: Foreward).

Implemented in 1990, the Texas Assessment of Academic Skills (TAAS) was designed to address the problems associated with the TEAMS. Like the TEAMS, the TAAS is a CRT which assesses reading, writing, and mathematics in Grades 3, 5, 7, 9, and 11. The TAAS also requires a written composition. Students are today required to pass the TAAS in order to be eligible to receive a high school diploma (Texas Education Agency, 1990: Foreward).

More than a test of basic skills, the TAAS design deemphasizes rote memorization, focusing instead on skills that "improve students' ability to think independently, read critically, write clearly, and solve problems logically" (Texas Education Agency, 1990: Foreward). In addition, its overall content encompasses more of the instructional targets of the State Board of Education Rules for Curriculum. Unlike the TEAMS, however, the TAAS measures only a portion of the instructional targets each year.

Because the targets selected for testing are not known in advance, instruction is more likely to include all of the instructional targets rather than simply what is tested. The result is a broader range of targets taught as well as tested (Texas Education Agency, 1990: Foreward).

In addition to the TAAS, the state enacted a norm-referenced testing program in April 1992. A norm-referenced test (NRT) is one which compares individual performance to the performance of a sample, or normative group. NRTs are generally normed at the national level. Test scores do not represent passing or failure, but instead reflect percentile rankings (Riverside, 1992: 5).

The Norm-Referenced Assessment Program for Texas (NAPT) was implemented to satisfy the state legislature's call for national normative data. The NAPT assessed students in Grades 3-11 in reading, language, mathematics, social studies, and science. Although the test does not include a diploma sanction, it serves to increase school accountability by comparing Texas students to their peers across the nation (Riverside, 1992: 5).

HIGH STAKES ASSESSMENT

It is common knowledge that the push for school accountability has created a high stakes climate for student assessment in Texas. Conversations with educators in the field and at TEA reflect shared concerns with the TAAS

diploma sanction, the TAAS passing standard, and accountability in education.

The most controversial aspect of the TAAS is the diploma sanction. Students and parents have trouble understanding why a student who has successfully completed all state-required coursework may be denied a high school diploma based on the results of a single test. As graduation approaches, emotions run high as students who have repeatedly failed to master the TAAS face the bleak prospect of not graduating with their classmates.

High stakes became even higher in 1991, the second year of the TAAS program. While the passing standard for the first year was 60%, the standard for the second year was raised to 70%. It is important to note that the State Board of Education initially adopted the 70% passing standard, and only allowed the 60% standard to help ease the transition from TAAS to TEAMS. Nonetheless, the higher standard was viewed by many as an arbitrary action which further reduced the ability of many students to graduate.

At the core of the high stakes climate is accountability. Test results are considered not only a measure of student achievement, but of teachers, schools, school districts, and the state's entire educational system. The renewal of contracts for teachers and administrators often hinges on the TAAS results. In addition, parents who are considering relocating frequently contact the Texas

Education Agency (TEA) to inquire about the performance of individual school districts.

PUBLIC REACTION TO STATEWIDE ASSESSMENT

Perhaps the most serious criticism of the state assessment program is that students are paying the price of the state's mistakes. The state has been criticized for raising assessment standards without improving the quality of instruction. Critics cite the fact that students entered school before the TAAS began, and complain that students who have been enrolled in state-approved below level courses are at an unfair disadvantage. It has been said that students are being penalized for poor teachers, administrators, and school boards as well as a lack of leadership by TEA (Cardenas, 1992: 17).

Legislators and taxpayers have also been blamed for their unwillingness to approve additional funding for educational programs (Cardenas, 1992: 17). Although funding for statewide student assessment has increased from practically nothing to \$13 million annually, bringing the overall amount spent on testing since 1978 to approximately \$40 million (Selby, 1992: 36), Texas still ranks 40th of the 50 states in total education dollars spent per pupil (Cardenas, 1992: 17).

Another serious charge against the TAAS program is that the test is in reality a measure of culture and class. Of the approximately 8,000 students denied high school diplomas

in May 1992 based on the TAAS results, nearly 73% are African American or Hispanic (Graves, 1992: 4). Critics of the TAAS contend that minorities fail the test in larger numbers than white students because minority students "are more likely to be culturally isolated and educationally and economically disadvantaged" (Evans, 1992: 15). In the spring of 1992, the Texas chapter of the National Association for the Advancement of Colored People (NAACP) asked its national leadership to work toward ending the TAAS graduation test. TEA denies that the test is biased, claiming that it merely reflects inequities in the state's funding of quality educational programs for school districts (Evans, 1992: 15).

The continual evolution of the statewide assessment program has also been criticized. School administrators complain that they are unable to compare results and measure growth because the tests are repeatedly changed. According to Ollie Besteiro, president of the Texas State Teachers Association, schools are simply "testing to get a score, to prove to some bureaucrat we are doing well" (Selby, 1992: 36). In addition, teachers and administrators claim that the emphasis on testing has detracted from other important school programs such as dropout prevention, drug abuse, and gang activity (Selby, 1992: 36).

Although the stakes involved in the NAPT program are much lower than those of the TAAS, the NAPT has nonetheless come under fire. Fed up with what they perceive to be

excessive testing, some parents refused to allow their children to participate in the NAPT. In Pasadena, Texas, a group of parents brought suit against TEA on charges that an NAPT test question asked students to make a judgement regarding different forms of religion. After reviewing the test, both a trial judge and an appellate judge determined that the test question had been interpreted out of context, and that it did not ask for a judgement call. Despite the judges' ruling against the parents, the suit generated a great deal of public resentment and negative publicity toward the statewide assessment program (Graves, 1992: 4).

STUDENT PERFORMANCE

Students in Texas public schools have been tested at the state level since 1980. So where do our students stand after all of these years? The results of the first TAAS, administered to 1.2 million students in Grades 3, 5, 7, 9, and 11 in October 1990, indicate the percentage of students who demonstrated mastery of the instructional targets of the state curriculum (Texas Education Agency, 1991: 2).

TAAS RESULTS - OCTOBER 1990

Grade 3:	64%
Grade 3 Spanish:	38%
Grade 5:	52%
Grade 7:	46%
Grade 9:	48%
Grade 11:	65%

On the first NAFT, administered to nearly 2.2 million students in Grades 3-11 in April 1992, the majority of Texas students scored above the national average. Test scores were generally above the national average at the earlier grades while slightly below the national average at the higher grade levels. The percentile rankings of Texas students are as follows (Texas Education Agency, July 1992: 5):

NAFT RESULTS - APRIL 1992

Grade 3:	61%
Grade 4:	55%
Grade 5:	59%
Grade 6:	53%
Grade 7:	47%
Grade 8:	49%
Grade 9:	49%
Grade 10:	53%
Grade 11:	48%

THE FUTURE OF TESTING IN TEXAS

The TAAS program is being redesigned, and a plan is underway to make the transition to the new assessment program. The goal is to expand the number of subject areas tested while simultaneously reducing the number of grade levels assessed. Beginning in the spring of 1993, students will be assessed at Grades 4, 8, and 10 in reading, writing, and mathematics (Texas Education Agency, April 1992: 1).

Assessment in science and social studies will be fully incorporated into the test battery by 1995. Computer

literacy assessment will also be required at the middle school level by 1995. End of course exams will be implemented at the secondary level in Algebra I, Biology I, Computer Science, and five additional subject areas between 1994 and 1995. Assessment in Physical Fitness/Health (elementary and middle school levels) and Oral Proficiency in a Second Language (elementary, middle school, and secondary levels) will be implemented by 1995. At this time, there are no plans to extend the diploma sanction beyond the reading, writing, and mathematics assessments (Texas Education Agency, April 1992: 1).

The NAEP will continue to be administered in Grades 3-11. Reading and mathematics assessment will be required in order to comply with reporting requirements for students participating in federally funded programs. However, assessment in language, social studies, and science will be optional (Texas Education Agency, April 1992: 1).

Perhaps the most significant change in state assessment is the move toward a primarily performance based assessment program. Performance assessment is considered an authentic measure of abilities not easily measured with conventional test questions. It is thought to allow students to demonstrate in-depth understanding and the integration of knowledge and abilities acquired in different areas. In addition, performance assessment is compatible with varied learning styles and multiple problem-solving approaches. It is hoped that performance assessment will improve the

quality of classroom instruction and improve student achievement (Texas Education Agency, summer 1992: ii).

This chapter presented an historical overview of the high stakes student assessment program in Texas. Student performance and public reaction to the testing program were described. Finally, the future of testing in Texas was discussed.

Chapter Four will address the methodology of the research. The survey design, sample selection, pretesting of the test instrument, and statistical methods used will be described. The strengths and weaknesses of survey research will also be discussed.

CHAPTER FOUR

INTRODUCTION

The purpose of this chapter is to describe the research methodology, and to demonstrate its utility in evaluating the attitudes of Texas educators toward multiple-choice testing and performance assessment. The discussion will focus on the sample selection process, the design of the survey instrument, and the statistical methods used to analyze the results.

METHODOLOGY

Data was collected from surveys mailed to school district test coordinators in Texas. Test coordinators are responsible for coordinating and supervising all standardized testing conducted within their school districts. They generally report to the superintendent of the school district. Test coordinators are required to hold college degrees in education and to have teaching experience.

Test coordinators were asked for their perceptions on a number of issues regarding multiple-choice testing and performance assessment. The survey was mailed July 13, 1992. Respondents were asked to return their surveys by August 1, 1992.

Since the general purpose of survey research is to measure attitudes and beliefs (Babbie, 1973: 57), it is

appropriate for measuring the attitudes of Texas educators toward two methods of student assessment. Although a case study could provide information regarding some factors involved in student assessment, such as the purposes for which test scores are used, a case study only allows for the study of a limited population. Survey research, on the other hand, provides an efficient and cost-effective method for measuring a large population. In addition, survey research can be used to determine whether districts are conducting locally designed performance assessments. Although a survey cannot indicate whether particular assessment issues have been overlooked, the use of a comments section allows for such input.

SAMPLE SELECTION

The sample drawn for the survey was a multiphase cluster sample. A list of all school districts in Texas (1060 total) was obtained from the Division of Instructional Outcomes Assessment at the Texas Education Agency. The list was sorted by the 20 education regions of the state. First, the 20 largest districts in terms of student population were selected. Second, the percentage of the total number of districts within the state was calculated for each region. In order to assure geographic representation, the number of districts selected from each region was based on the region's percentage of the total number of districts in the state. Finally, random selection from each region was

conducted. The total number selected for the sample was 130. The surveys were mailed to district test coordinators.

A cover letter introduced the author, and explained the purpose of the survey. A copy of the cover letter is located in Appendix C. The survey is located in Appendix B.

SURVEY DESIGN

Organization

The survey asked for responses regarding six broad categories of issues which impact attitudes toward multiple-choice testing and performance assessment. These categories represent expert opinion as derived from a review of the literature on student assessment. The categories are: test design; impact on teaching and curriculum; impact on students; equity in testing; cost; and time. Underlying facets within each category were addressed.

Because no single method for recording responses was appropriate for all of the survey questions, the survey was divided into four sections of questions. The first two sections asked for responses to identical statements regarding first, multiple-choice testing, and second, performance assessment. A Likert scale was employed which used Strongly Agree (SA), Agree (A), No Opinion (NO), Disagree (D), and Strongly Disagree (SD) as the answer choices.

The third section of questions explicitly compared multiple-choice testing to performance assessment.

Participants were asked to respond to questions using yes/no responses which were later coded through percentage distributions.

The fourth set of questions was designed to obtain demographic information to be coded through percentage distributions. The final section provided for open-ended comments.

Pre-test

The survey was pre-tested at the Texas Education Agency to obtain preliminary feedback. Ten professional staff members with expertise in teaching and student assessment were asked to complete the survey. Their input provided valuable information in both clarifying the survey questions and determining their location within the survey.

CATEGORIES OF ISSUES

Test design

The first category of issues addressed the underlying facets of test design which impact attitudes toward student assessment. These include skills measured, range and focus of coverage, flexibility for demonstrating achievement, and scoring.

Although multiple-choice testing offers a broad range of coverage and objective scoring, it has been criticized for its shallow focus of coverage and its emphasis on basic

skills. In addition, multiple-choice tests offer students little flexibility for demonstrating their academic achievement.

By comparison, performance assessment offers a narrower range of coverage, but provides a more in-depth focus. Performance assessment is thought to measure higher-order thinking skills in a manner that allows for increased flexibility in the demonstration of academic achievement. It is, however, criticized for its potential for subjectivity in scoring.

Impact on Teaching and Curriculum

This category addressed the impact of each testing method on teaching and curriculum. The types of skills taught, teaching methods, curriculum content coverage, and textbook content coverage were emphasized.

One of the most significant criticisms of multiple-choice testing is that it drives teachers to concentrate on teaching basic skills, and to limit their teaching methods to worksheets and lectures. Critics of multiple-choice testing argue that the scope of coverage in curriculum and textbooks has narrowed as a result of the tests' emphasis on basic skills.

In contrast, supporters of performance assessment contend that its emphasis on higher-order thinking skills will encourage teachers to focus on these skills by offering students opportunities to write, research, discuss, and

demonstrate. It is hoped that the scope of curriculum and textbook coverage will be broadened as a result.

Impact on Students

This category focused on the impact of each testing method on students. The issues included are program placement, promotion/retention, and whether the assessment is meaningful to students.

Schools rely extensively on standardized test results to place students in instructional programs and to determine whether students are promoted or retained in grade. Critics charge that multiple-choice tests produce trivial and superficial diagnostic information, and that they are an artificial predictor of student ability. Those who advocate performance assessment believe that it offers students a meaningful experience which provides in-depth information that is truly representative of student achievement.

Equity in Testing

This category involved issues related to equity in testing. The focus was on minority students, low-income students, limited English proficient (LEP) students, and the allowance of each testing method for varied learning styles.

Multiple-choice tests have been labelled as culturally biased because they reflect the lifestyles and experiences of white, middle class students. Furthermore, their allowance for varied learning styles is minimal.

In contrast, performance assessment clearly allows for the varied learning styles of students. The equity of performance assessment in its measurement of minority, low-income, and LEP students, however, is as yet undetermined.

Cost

The questions in this section pertained to cost issues in testing. The costs associated with test development, equipment/materials, and scoring were addressed.

Multiple-choice testing is unarguably more cost-efficient than performance assessment in the areas of test development, equipment/materials requirements, and scoring. Multiple-choice tests have been developed in abundance. The equipment and materials requirements for schools are limited to the provision of No. 2 pencils. In addition, computer scoring is efficient and inexpensive.

On the other hand, performance assessments are more complicated and therefore more costly to develop. Performance assessments may also require the use of laboratory equipment or reference materials. Furthermore, hand scoring represents a significant expense.

Time

This category addressed time-related issues. These questions pertain to test development, test administration, and scoring.

The format of multiple-choice tests allows for relatively rapid test development and test administration. In addition, computerized scoring ensures that test results are quickly produced and reported.

Performance assessments, however, are time-consuming to develop. Test administration is also more time-consuming because teachers and students are less familiar with the format, and the level of teacher involvement is therefore increased. Furthermore, assessments which require the use of libraries or laboratories add to the time required for testing. Finally, the need for performance assessments to be hand scored lengthens the scoring process.

ISSUES/SURVEY QUESTIONS

The six categories of issues and the survey questions which correspond to each category are displayed in Table 4 on the following page.

ISSUES/SURVEY QUESTIONS

Table 4

ISSUES CATEGORY	ADDRESSED IN QUESTIONS
Test Design	MC #1 PA #1 C #1, 2, 3
Impact on Teaching & Curriculum	MC #2, 3, 4 PA #2, 3, 4 C #4
Impact on Students	MC #5, 6, 7 PA #5, 6, 7
Equity in Testing	MC #8, 9, 10, 11 PA #8, 9, 10, 11
Cost	C #5, 6, 7
Time	C #8, 9, 10

MC = Multiple-Choice testing

PA = Performance Assessment

C = Comparative Questions

STATISTICAL METHODS

A Chi-square test was employed to test for differences in attitudes and perceptions on questions which used a Likert scale (sections one and two). The Chi-square results reveal whether test coordinators found differences between multiple-choice and performance based testing instruments. The Chi-square formula was obtained from Roger Kirk's textbook on statistical methods (Kirk, 1984). Percentage

distributions were used to analyze yes/no responses (section three).

Demographic questions (section four) were measured through percentage distributions. The final section of the survey allowed for open-ended responses.

STRENGTHS AND WEAKNESSES OF SURVEY RESEARCH

Perhaps the greatest strength of survey research is that it allows for the measurement of a large population. In addition, its flexibility allows many topics to be addressed. Survey research is also strong on reliability. Finally, the standardization of the survey allows for the development of generalizations.

One weakness of survey research is that because the design of the instrument cannot be altered once implemented, flexibility is limited. Another weakness is that the standardization of survey research may yield superficial data. Furthermore, the use of predetermined responses is artificial in its failure to allow for other responses. Finally, this artificiality has a negative impact on validity.

This chapter described the general characteristics of survey research. Also discussed were the sample selection process, survey design, and statistical methods employed in the research. The strengths and weaknesses of survey research were outlined.

In Chapter Five, the survey results will be discussed. The results are organized according to the six broad categories of issues: test design; impact on teaching and curriculum; impact on students; equity in testing; cost; and time.

CHAPTER FIVE

RESULTS

This chapter will discuss the results of the survey sent to school district test coordinators around Texas. Responses will be discussed according to categories of issues as follows: test design; impact on teaching and curriculum; impact on students; equity in testing; cost; and time. The respondents' attitudes toward multiple-choice testing and performance assessment, as well as areas of particular concern, will be evident. This chapter will demonstrate how the attitudes of Texas educators compare to those of educators around the nation as discussed in the literature review in Chapter Two.

The survey questions did not require test coordinators to choose between multiple-choice testing and performance assessment; instead, respondents were asked to evaluate each method in accordance with the categories of issues. It was therefore possible for an individual to find both methods appropriate for a given purpose.

One hundred thirty surveys were mailed. Fifty eight were returned, representing a return rate of 44.6 percent. The results listed here represent responses of strongly agree and agree for those questions which used a Likert scale, and yes for questions which used a yes/no response.

A copy of the survey, including the number of responses to each question and all raw data, is located in Appendix B.

TEST DESIGN

SKILLS
Table 5.1

QUESTION	MC % A/SA	PA % A/SA	CHI- SQUARE
effectively measures skills required in the real world	36.3 (55)	76.7 (56)	23.5*

* significant at the .001 level
() = # of respondents
MC = multiple-choice
PA = performance assessment
A/SA = agree or strongly agree

Seventy six percent of the district test coordinators agreed or strongly agreed that performance assessment effectively measures skills required in the real world. Slightly over one third of the respondents felt similarly about multiple-choice testing. The difference was significant at the .001 level (see Table 5.1). This finding corresponds with the literature because multiple-choice testing has been criticized for focusing on basic skills while excluding higher-order thinking skills. In contrast, performance assessment emphasizes independent thought, critical analysis, writing, and problem solving.

TEST DESIGN

COVERAGE/FLEXIBILITY/SCORING
Table 5.2

QUESTION	YES
depth of coverage is more important than breadth	54.7% (53)
PA offers flexibility in demonstration of academic achievement	91.2% (57)
PA can achieve acceptable scoring objectivity and uniformity	50.9% (51)

() = # of respondents

PA = performance assessment

Table 5.2 reveals an overwhelming preference for performance assessment in the area of flexibility for demonstrating achievement. In this regard, performance assessment was preferred to multiple-choice testing by 91 percent of the respondents. This preference reflects today's awareness that students learn in different ways, and that a student who performs poorly in one format may demonstrate understanding through the use of a different approach.

Test coordinators were asked whether they consider depth of coverage more important than breadth of coverage. Opinion was evenly divided. While multiple-choice testing has a broad range and shallow focus, performance assessment offers a narrower range with more in-depth coverage. Supporters of multiple-choice testing claim that a broad

range is important because of differing opinions on what should be taught. Advocates of performance assessment cite the futility of trying to teach everything of importance, favoring instead the in-depth assessment of those skills considered most important.

Opinion was also evenly mixed on objectivity and uniformity in scoring. While the machine scoring of multiple-choice tests offers a strong guarantee of objectivity and uniformity, performance assessments are hand scored. Approximately half of the test coordinators felt that performance assessment cannot achieve acceptable levels of objectivity and uniformity in scoring.

IMPACT ON TEACHING AND CURRICULUM

FOCUS/METHODS/CONTENT
Table 5.3

QUESTION	MC % A/SA	PA % A/SA	CHI- SQUARE
encourages teaching that focuses on skills required as adults	36.8 (57)	73.6 (57)	28.8*
promotes use of variety of teaching methods	29.3 (58)	74.5 (55)	32*
promotes curriculum that is broad and balanced	37.9 (58)	67.8 (56)	24.8*

* significant at the .001 level

() = # of respondents

MC = multiple-choice

PA = performance assessment

A/SA = agree or strongly agree

The respondents maintained that performance assessment has a favorable impact on teaching and curriculum. All items revealed differences in attitudes toward multiple-choice testing and performance assessment that were significant at the .001 level (see Table 5.3).

Nearly three-fourths of the test coordinators felt that performance assessment encourages teachers to focus on the kinds of skills students will require as adults. Only one-third agreed that multiple-choice testing has this effect. This is an important consideration because prospective employers complain that high school graduates often lack the ability to analyze and solve problems, draw appropriate conclusions, and work cooperatively.

Table 5.3 also indicates that respondents felt that teaching methods reflect the assessment process. Nearly three-fourths agreed that performance assessment encourages teachers to employ a variety of teaching methods.

It is evident that educators believe that performance assessment is more likely to promote curriculum that is broad and balanced in content. Almost 70 percent of the respondents considered performance assessment effective in this regard. One strength of performance assessment is that it generally requires the integration of knowledge because it measures not only a variety of skills, but also measures subject knowledge in more than one area at the same time. Less than 40 percent of those who responded gave a favorable rating to multiple-choice testing in this area.

IMPACT ON TEACHING AND CURRICULUM

TEXTBOOKS Table 5.4

QUESTION	YES
PA likely to have a narrowing effect on textbook content coverage	48% (52)

() = # of respondents

PA = performance assessment

The perceived effect of assessment on textbooks is revealed in Table 5.4. Multiple-choice tests have been criticized for having a narrowing effect on textbook content coverage. Interestingly, performance assessment did not fare well in this regard. Forty eight percent of the respondents thought that performance assessment would have a similar narrowing effect. It is interesting to note that the respondents felt that performance assessment would have a broadening effect on curriculum without having a similar impact on textbook content coverage.

IMPACT ON STUDENTS

PLACEMENT/MEANINGFULNESS
Table 5.5

QUESTION	MC % A/SA	PA % A/SA	CHI- SQUARE
provides valuable info. for decisions regarding student program placement	66 (56)	73.2 (56)	5.2
provides valuable info. for decisions regarding student promotion and retention	57.1 (56)	71.4 (56)	11.7*
offers students meaningful educational experience	17.2 (58)	68.4 (57)	37.3**

* significant at the .05 level

** significant at the .001 level

() = # of respondents

MC = multiple-choice

PA = performance assessment

A/SA = agree or strongly agree

The results regarding the impact of each testing method on students are displayed in Table 5.5. Test coordinators felt that both testing methods provide information useful for making decisions regarding student program placement, promotion, and retention. Performance assessment, however, was considered a more meaningful experience for students by an overwhelming majority of the respondents.

It is noteworthy that although the respondents strongly supported the use of performance assessment for making decisions regarding student program placement and promotion/retention, multiple-choice testing also received

relatively strong support. Fifty seven percent of the test coordinators supported multiple-choice testing in the area of promotion/retention as compared to 71 percent for performance assessment. The difference was significant at the .05 level. However, there was no significant difference between the approval of multiple-choice testing and performance assessment in the area of program placement.

The respondents overwhelmingly perceived performance assessment as superior to multiple-choice testing in the ability to offer students a meaningful educational experience (68 percent for performance assessment versus 17 percent for multiple-choice testing). In this area, the difference in attitudes toward the two testing methods was significant at the .001 level. Despite the perceived value of multiple-choice testing as a decision-making tool, the respondents apparently considered it meaningless to students.

EQUITY IN TESTING

ETHNICITY/INCOME/LANGUAGE/LEARNING STYLES
Table 5.6

QUESTION	MC % A/SA	PA % A/SA	CHI- SQUARE
equitable measure of minority students	31.5 (57)	62.5 (56)	15.5*
equitable measure of low-income students	33.3 (57)	64.2 (56)	14*
well-suited to measure LEP students	10.7 (56)	54.7 (53)	37.6**
well-suited for varied learning styles	9 (55)	67.2 (55)	43.3**

* significant at the .01 level

** significant at the .001 level

() = # of respondents

MC = multiple-choice

PA = performance assessment

A/SA = agree or strongly agree

The results for questions regarding equity in testing may be observed in Table 5.6. Although the equity of test instruments has improved in recent years, the respondents clearly had serious concerns about the equity of multiple-choice tests as a measure of the ability of minority and low-income students. Test coordinators were particularly concerned about the equity of multiple-choice tests in assessing limited English proficient (LEP) students and in addressing the varied learning styles of students. The respondents were obviously more comfortable with performance assessment with regard to equity in testing. The

differences between the two testing methods for all questions in this category were significant at the .01 level and above.

Performance assessment was approved by a 2-1 margin over multiple-choice testing as an equitable measure of minority and low-income students. For the measurement of LEP students, multiple-choice testing received an extremely low approval rate of 10 percent, and it is interesting to note that performance assessment received less support than it received in other areas addressing test equity. Only 54 percent of those who responded agreed that performance assessment is an equitable measure of the academic achievement of LEP students.

Obviously, LEP students face unique problems with each method of testing. Although performance assessment may be more equitable in selected areas, such as conducting science experiments or performing mathematical calculations, the language barrier may render performance assessment little better than multiple-choice testing in assessments that require reading, writing, and research.

Less than 10 percent of the respondents agreed that multiple-choice testing is well-suited to accommodate varied learning styles. In contrast, two-thirds of the respondents felt that performance assessment accommodates a variety of learning styles. Performance assessment, like real life, allows for variation in the approach to tasks and topics.

COST

BENEFITS/MATERIALS
Table 5.7

QUESTION	YES
benefits of PA outweigh expense	45.4% (55)
additional materials/equipment for PA poses financial problem for school districts	81% (58)
depth of info. provided by PA justifies cost	40% (50)

() = # of respondents

PA = performance assessment

The survey provided an overall cost comparison of multiple-choice testing and performance assessment, and explained that the additional cost of performance assessment is largely due to the cost of test development and scoring. The results displayed in Table 5.7 indicate that support for performance assessment is greatly reduced by its higher cost.

It is obvious that the respondents considered performance assessment too expensive. Only 45 percent agreed that the general benefits of performance assessment outweigh its expense. Despite the fact that nearly three-fourths of the respondents agreed that performance assessment provides valuable information for placing, promoting, and retaining students, only 40 percent felt that

the depth of the information provided justifies its overall cost.

While the equipment and materials requirements for multiple-choice testing are minimal, performance assessment may require testing students in libraries or science laboratories. Additional materials may be required. Table 5.7 reveals that 81 percent of the respondents felt that accommodating these additional requirements poses a significant problem for school districts.

TIME

DEVELOPMENT/ADMINISTRATION/SCORING
Table 5.8

QUESTION	YES
complexity of test development renders statewide PA impractical	64.7% (51)
instructional value of PA justifies lost instructional time	41.1% (51)
increased time for scoring PA poses major problem for school districts	75.4% (57)

() = # of respondents
PA = performance assessment

As shown in Table 5.8, most respondents felt that performance assessment is too time-consuming. Sixty four percent responded that performance assessment at the state level is impractical due to the time-consuming and complex process of test development.

Considering the pressure on schools to improve student performance, it is no surprise that the amount of time consumed by testing was an important issue. Only forty one percent of the test coordinators felt that the instructional value of performance assessment is sufficient to justify the increased time it requires.

The amount of time required for reporting test results is critical because student remediation, promotion, and retention are so closely linked to test results. Performance assessment fared poorly in this area. Fully three-fourths of the district test coordinators felt that the time-consuming scoring of performance assessments poses significant problems for school districts.

DEMOGRAPHIC INFORMATION

EXPERIENCE Table 5.9a

YEARS SERVED AS TEST COORD.

	0-5	6-10	11-20	>20
responses of test coordinators (58)	32.7%	36.2%	31%	0%

() = # respondents

STUDENT POPULATIONS

Table 5.9b

% OF MINORITY/LOW-INCOME/
LEP STUDENTS IN DISTRICT

0-25% 26-50% 51-75% >75%

responses of test coordinators (57)	21%	40.3%	28%	10.5%
---	-----	-------	-----	-------

() = # of respondents

USAGE OF TEST RESULTS

Table 5.10a

% R % O % F

district uses test scores as sole basis for decisions on student program placement, promotion, retention (57)	61.4	33.3	5.2
--	------	------	-----

R = rarely

O = occasionally

F = frequently

() = # of respondents

LOCAL ASSESSMENT

Table 5.10b

% R % O % F

district conducts locally designed PA (57)	38.5	54.3	7
--	------	------	---

R = rarely

O = occasionally

F = frequently

() = # of respondents

PA = performance assessment

Tables 5.9a and 5.9b display demographic information regarding the respondents and student populations. Table

5.10a addresses the uses of test results in school districts.

Approximately 60 percent of the respondents claimed that standardized test scores are rarely the sole basis in their districts for making decisions regarding student program placement, promotion, and retention. One-third indicated that test scores are occasionally the sole basis for these decisions. While the 60 percent figure should encourage those who criticize the "tracking" of students based on test scores, it would be interesting to know which circumstances allow for the use of test results as the sole basis for decision making.

Table 5.10b reveals that approximately half of the respondents stated that their districts conduct locally designed performance assessments. A comparison of locally designed assessments to those developed by the state would be of interest.

This chapter presented the results of the survey of Texas school district test coordinators. The results were organized according to the categories of test design, impact on teaching and curriculum, impact on students, equity in testing, cost, and time. Demographic information regarding the respondents and their school districts was also provided.

The final chapter will discuss the conclusions drawn from the research. The advantages and disadvantages of

multiple-choice testing and performance assessment will be summarized.

CHAPTER SIX

CONCLUSION

This chapter summarizes the attitudes of school district test coordinators toward multiple-choice testing and performance assessment. The discussion focuses on the six broad categories of issues that were addressed in the survey.

TEST DESIGN

School district test coordinators preferred performance assessment to multiple-choice testing by a 2-1 margin as a tool for effectively measuring the skills required in the real world. They also demonstrated overwhelming support for performance assessment in the flexibility it allows students for demonstrating academic achievement.

Breadth versus depth of test content coverage, however, was not a significant issue to test coordinators. In addition, the fact that the scoring of multiple-choice tests is generally more objective than that of performance assessments did not appear to pose a major concern to the respondents.

IMPACT ON TEACHING AND CURRICULUM

Test coordinators obviously believed that performance assessment has a more beneficial effect on teaching and curriculum than multiple-choice testing. They agreed that

CHAPTER SIX

CONCLUSION

This chapter summarizes the attitudes of school district test coordinators toward multiple-choice testing and performance assessment. The discussion focuses on the six broad categories of issues that were addressed in the survey.

TEST DESIGN

School district test coordinators preferred performance assessment to multiple-choice testing by a 2-1 margin as a tool for effectively measuring the skills required in the real world. They also demonstrated overwhelming support for performance assessment in the flexibility it allows students for demonstrating academic achievement.

Breadth versus depth of test content coverage, however, was not a significant issue to test coordinators. In addition, the fact that the scoring of multiple-choice tests is generally more objective than that of performance assessments did not appear to pose a major concern to the respondents.

IMPACT ON TEACHING AND CURRICULUM

Test coordinators obviously believed that performance assessment has a more beneficial effect on teaching and curriculum than multiple-choice testing. They agreed that

performance assessment supports efforts to teach higher-order thinking skills through a variety of classroom teaching methods. Their perception of a positive relationship between performance assessment and curriculum was also evident. Nonetheless, test coordinators did not feel that the positive effects of performance assessment would extend to textbooks.

The respondents obviously had a negative perception of multiple-choice testing in its impact on teaching and curriculum. Those who believed that multiple-choice testing promotes the teaching of higher-order thinking skills and has a broadening effect on teaching and curriculum constituted only about one-third of the test coordinators.

IMPACT ON STUDENTS

Test coordinators believed that multiple-choice testing and performance assessment are generally equal in their usefulness for making student program placement decisions. They indicated, however, that student promotion and retention decisions are better served through the use of performance assessment.

The meaningfulness of assessment was an important consideration to the respondents. The perception of the superiority of performance assessment in this regard was undeniable.

EQUITY IN TESTING

The equity of multiple-choice testing in the assessment of minority and low-income students was of great concern to test coordinators. Furthermore, they were overwhelmingly critical of multiple-choice testing in the measurement of students with limited English proficiency (LEP). In contrast, test coordinators showed strong support for the equity of performance assessment toward minority and low-income students. Their support for performance assessment, however, was significantly weaker in the measurement of LEP students.

In this category of issues, the most obvious disparity between attitudes on multiple-choice testing and performance assessment was evidenced in the suitability of each method for accommodating varied learning styles. Test coordinators preferred performance assessment by a margin of 7-1.

COST

Test coordinators reported serious reservations regarding the costs associated with performance assessment. The majority of the respondents felt that the higher cost of performance assessment outweighs the benefits it provides. Test coordinators were most concerned with the financial burden for school districts in providing the additional materials and equipment often required by performance assessment.

TIME

Of the six categories of issues addressed by the survey, test coordinators were most concerned with the amount of time consumed by performance assessment. The majority of the respondents considered the complexity of test development a major barrier to statewide performance assessment. Most agreed that lost instructional time is more important than the instructional value of performance assessment. In addition, the lengthier scoring process was considered the most serious drawback of the time-related issues.

SUMMARY

There are obviously no easy solutions for selecting a method of assessment that is appropriate for all students. Clearly there are advantages and disadvantages associated with both multiple-choice testing and performance assessment. What emerges is a series of trade-offs that are inherent in the selection of either testing method.

The respondents maintained that multiple-choice testing offers a cost effective and objective approach to the assessment of large groups of students in a timely manner. A broad range of subject knowledge may be measured with a minimum of school materials and equipment. The trade-off is that the design of multiple-choice tests restricts both the kinds of skills that may be assessed and the depth of the

information provided. Furthermore, teaching and curriculum tend to mirror this narrowed approach. The assessment process is therefore often a meaningless experience for students, and may also be inequitable.

Performance assessment offers students assessment in a meaningful context that better targets the types of skills required in the real world. As a result, teaching methods and curriculum are enhanced by instructionally sound assessment. As an in-depth measure of academic achievement, the prospects for equity in testing are improved. The design of performance assessment, however, limits the breadth of subject coverage and the objectivity of test results. In addition, performance assessment is considerably more time-consuming and expensive to conduct.

The best assessment at this time is likely one which incorporates the best of both methods. The statewide assessment program in Texas is moving in this direction as state education policymakers begin the incorporation of extensive performance assessment into the existing multiple-choice testing program. It appears that state policymakers recognize that there is a place for both multiple-choice testing and performance assessment in a sound assessment program.

APPENDIX ATABLE 1
TEST DESIGN

<u>FACET/EXPERT OPINION</u>	<u>MULTIPLE-CHOICE</u>	<u>PERFORMANCE ASSESSMENT</u>
skills measured (U.C.L.A.; Wiggins)	basic	higher-order thinking
range/focus of coverage (Arter; Feinberg; Linn et al.)	broad/shallow	narrow/in-depth
flexibility for demonstrating achievement (Wiggins)	minimal	substantial
scoring (Feinberg)	objective	potentially subjective

TABLE 2
IMPACT ON TEACHING AND CURRICULUM

<u>FACET/EXPERT OPINION</u>	<u>MULTIPLE-CHOICE</u>	<u>PERFORMANCE ASSESSMENT</u>
skills taught (Haney & Madaus)	basic	higher-order thinking
teaching methods (Darling-Hammond; Wiggins)	worksheets/ lectures	writing/research/ discussion/ demonstration
curriculum content coverage (Airasian)	narrowed	broadened
textbook content coverage (Neill & Medina)	narrowed	broadened

TABLE 3
IMPACT ON STUDENTS

<u>FACET/EXPERT OPINION</u>	<u>MULTIPLE-CHOICE</u>	<u>PERFORMANCE ASSESSMENT</u>
program placement (Darling-Hammond; Neill & Medina)	superficial info. to aid diagnosis	in-depth info. to aid diagnosis
promotion/retention (Airasian; Neill & Medina)	artificial predictor of ability	authentic predictor of ability
meaningful (Archbald & Newmann; Linn et al.; Shepard)	trivial/ meaningless	meaningful

TABLE 4
EQUITY IN TESTING

<u>FACET/EXPERT OPINION</u>	<u>MULTIPLE-CHOICE</u>	<u>PERFORMANCE ASSESSMENT</u>
minority students (Darling-Hammond; Neill & Medina)	questionable	undetermined
low-income students (Darling-Hammond; Neill & Medina)	questionable	undetermined
limited English proficiency (Ascher; Neill & Medina)	questionable	undetermined
allowance for varied learning styles (Neill & Medina)	minimal	extensive

TABLE 5
COST

<u>FACET/EXPERT OPINION</u>	<u>MULTIPLE-CHOICE</u>	<u>PERFORMANCE ASSESSMENT</u>
development (Cizek; Mitchell)	inexpensive	expensive
equipment/materials (U.C.L.A.)	inexpensive	potentially expensive
scoring (Mitchell; U.C.L.A.)	inexpensive	expensive

TABLE 6
TIME

<u>FACET/EXPERT OPINION</u>	<u>MULTIPLE-CHOICE</u>	<u>PERFORMANCE ASSESSMENT</u>
development (U.C.L.A.)	quickly developed	time-consuming
test administration (Archbald & Newmann; U.C.L.A.)	quickly administered	potentially time-consuming
scoring (Archbald & Newmann; U.C.L.A.)	quickly scored	time-consuming

APPENDIX B**STATEWIDE PERFORMANCE ASSESSMENT IN TEXAS: A SURVEY OF THE
ATTITUDES OF SCHOOL DISTRICT TEST COORDINATORS**

This survey reflects a consensus of expert opinion regarding student assessment. It has been designed to assess the attitudes of district test coordinators regarding performance assessment at the state level.

The survey is divided into three sets of questions. The first set pertains to multiple-choice testing; the second pertains to performance assessment; the third set of questions explicitly compares the two types of assessment.

Please indicate your response by circling the appropriate answer.

SA = Strongly Agree

D = Disagree

A = Agree

SD = Strongly Disagree

NO = No Opinion

MULTIPLE-CHOICE TESTING:

- effectively measures skills required in the real world.

SA	A	NO	D	SD
2	18	5	22	8
total 55				

- encourages teaching that focuses on the kinds of skills students will require as adults.

SA	A	NO	D	SD
1	20	2	24	10
total 57				

- promotes the use of a variety of classroom teaching methods.

SA	A	NO	D	SD
1	16	8	22	11
total 58				

- promotes school curriculum which is broad and balanced in content.

SA	A	NO	D	SD
4	18	5	20	11
total 58				

MULTIPLE-CHOICE TESTING (continued):

5. provides valuable diagnostic information for making decisions regarding student program placement.

SA	A	NO	D	SD
3	34	5	9	5
total 56				

6. provides valuable information for making decisions regarding student promotion and retention.

SA	A	NO	D	SD
1	31	3	13	8
total 56				

7. offers students a meaningful educational experience.

SA	A	NO	D	SD
1	9	17	22	9
total 58				

8. is an equitable measure of the academic achievement of minority students.

SA	A	NO	D	SD
2	16	10	20	9
total 57				

9. is an equitable measure of the academic achievement of low-income students.

SA	A	NO	D	SD
2	17	10	19	9
total 57				

10. is well-suited to measure the academic achievement of limited English proficient students.

SA	A	NO	D	SD
0	6	8	29	13
total 56				

11. is well-suited to accommodate the varied learning styles of students.

SA	A	NO	D	SD
0	5	5	31	14
total 55				

PERFORMANCE ASSESSMENT:

1. effectively measures skills required in the real world.

SA	A	NO	D	SD
7	36	6	5	2
total 56				

2. encourages teaching that focuses on the kinds of skills students will require as adults.

SA	A	NO	D	SD
6	36	7	7	1
total 57				

3. promotes the use of a variety of classroom teaching methods.

SA	A	NO	D	SD
6	35	8	6	0
total 55				

4. promotes school curriculum which is broad and balanced in content.

SA	A	NO	D	SD
8	30	11	7	0
total 56				

5. provides valuable diagnostic information for making decisions regarding student program placement.

SA	A	NO	D	SD
6	35	8	6	1
total 56				

6. provides valuable information for making decisions regarding student promotion and retention.

SA	A	NO	D	SD
5	35	7	8	1
total 56				

7. offers students a meaningful educational experience.

SA	A	NO	D	SD
8	31	12	5	1
total 57				

PERFORMANCE ASSESSMENT (continued):

8. is an equitable measure of the academic achievement of minority students.

SA	A	NO	D	SD
3	32	9	11	1
total 56				

9. is an equitable measure of the academic achievement of low-income students.

SA	A	NO	D	SD
3	33	9	9	2
total 56				

10. is well-suited to measure the academic achievement of limited English proficient students.

SA	A	NO	D	SD
3	26	13	9	2
total 53				

11. is well-suited to accommodate the varied learning styles of students.

SA	A	NO	D	SD
9	28	5	10	3
total 55				

COMPARATIVE QUESTIONS

1. While multiple-choice testing offers a broad range of coverage, performance assessment offers an in-depth focus on student achievement. Would you say that depth of coverage is more important than breadth of coverage?

yes	no
29	24
total 53	

2. Do you feel that performance assessment offers students a flexibility in the demonstration of academic achievement that is not possible with multiple-choice testing?

yes	no
52	5
total 57	

COMPARATIVE QUESTIONS (continued):

3. Supporters of multiple-choice testing claim that the potential for subjectivity in scoring is a major problem in performance assessment. Do you think that performance assessment can achieve acceptable levels of objectivity and uniformity in scoring?

yes	no
26	25

total 51

4. Some experts claim that undue reliance on multiple-choice testing has caused a narrowing of curriculum which, in turn, has resulted in the narrowing of textbook content coverage. Assuming this to be true, would you say that performance assessment is likely to have a similar impact on curriculum and textbooks?

yes	no
25	27

total 52

5. Performance assessments are costly to develop. Given that some states have reported the overall per student cost of multiple-choice testing to be \$1.50 as compared to \$3-10 for performance assessment, do you feel that the benefits of performance assessment outweigh the expense?

yes	no
25	30

total 55

6. Multiple-choice tests require only that students be supplied with No. 2 pencils while performance assessment may require the use of libraries and laboratory equipment. Do you think that accommodating the additional requirements of performance assessment within existing facilities poses a significant financial problem for school districts?

yes	no
47	11

total 58

7. It is estimated that the overall cost of performance assessment is 3-7 times greater than that of multiple-choice testing. Scoring represents a significant portion of the increased cost of performance assessment. Would you say that the depth of the information provided by performance assessment justifies the cost?

yes	no
20	30

total 50

COMPARATIVE QUESTIONS (continued):

8. Multiple-choice tests are more quickly developed than performance assessments. In your opinion, does the complexity of test development render statewide performance assessment impractical?

	yes	no
	33	18
total	51	

9. Performance assessments are more time-consuming to administer than multiple-choice tests. Do you think that the instructional value of performance assessment justifies the lost instructional time?

	yes	no
	21	30
total	51	

10. Do you feel that the increased time required for scoring performance assessments as compared to multiple-choice tests poses a major problem for school districts?

	yes	no
	43	14
total	57	

ADDITIONAL INFORMATION

1. Please indicate the number of years you have served as a district test coordinator.

0-5	6-10	11-20	more than 20
19	21	18	0
total	58		

2. Please estimate the percentage of students in your school district who are minority, low-income, and/or limited English proficient students.

0-25	26-50	51-75	more than 75
12	23	16	6
total	57		

3. To what extent are standardized test scores the sole basis for decisions regarding student program placement, promotion, and retention in your district?

rarely	occasionally	frequently
35	19	3
total	57	

ADDITIONAL INFORMATION (continued):

4. To what extent do schools in your district conduct locally designed performance assessments?

	rarely	occasionally	frequently
	22	31	4
total	57		

COMMENTS

Thank you for your participation in this survey.

APPENDIX C

July 13, 1992

Dear District Coordinator:

My name is Jana Curtis. I am a graduate student at Southwest Texas State University. One of my final requirements for graduation is to complete an Applied Research Project in which I am surveying a randomly selected sample of school district test coordinators across Texas.

Current debate regarding the shortcomings of standardized testing has spawned national interest in performance assessment. The purpose of my research is to determine the attitudes of Texas educators toward performance assessment at the state level. My research will also compare traditional multiple-choice testing to performance assessment. As a district test coordinator, you are uniquely qualified to address the issues involved in testing students.

The questions on the attached survey reflect a consensus of expert opinion derived from a review of the current literature on student assessment. The survey will only require a few minutes to complete. This research is strictly for my educational purposes, and all responses will remain anonymous. If you wish to have a copy of the results of my survey, please write your name and address on the survey. I will be happy to provide you with my results.

Please complete and mail the survey by August 1. I have enclosed a stamped self-addressed envelope for your convenience. If you have any questions about my research, I may be reached at home at (512) 441-8616. Your reply will be greatly appreciated!

Sincerely,

Jana Curtis
1406 Folts Avenue
Austin, TX 78704
(512) 441-8616

BIBLIOGRAPHY

- Airasian, Peter W. (1987). State mandated testing and educational reform: context and consequences. American Journal of Education, 95, 393-412.
- Archbald, Doug A., & Newmann, Fred M. (1988). Beyond standardized testing: assessing authentic academic achievement in the secondary school. Madison: University of Wisconsin, National Center on Effective Secondary Schools.
- Arter, Judy. (1991). Performance assessment: what's out there and how useful is it really? Annual Meeting of the American Educational Research Association (pp. 2-7). Washington, D.C.: Office of Educational Research and Improvement.
- Ascher, Carol. (1991). Testing bilingual students: do we speak the same language? PTA Today, 16, 7-9.
- Ashworth, Kenneth H. (1989/1990). Standardized testing: a defense. The College Board Review, 10, 23-24. (From The Education Digest, 1990, 56).
- Babbie, Earl R. (1973). Survey research methods. Belmont, CA: Wadsworth.
- Babbie, Earl. (1989). The practice of social research (5th ed.). Belmont, CA: Wadsworth.
- Berdic, Douglas R., & Anderson, John F. (1974). Questionnaires: design and use. Metuchen, NJ: Scarecrow Press.

- Brescia, William, & Fortune, Jim C. (1988). Standardized testing of American Indian students. (Report No. EDO-RC-88-15). Las Cruces, NM: ERIC Clearinghouse on Rural Education in Small Schools. (ERIC Document Reproduction Service No. ED 296 813).
- Cardenas, Jose A. (1992, June). Students are being punished for states' mistakes [Letter to the editor]. Austin American-Statesman, p. 17.
- Center for Research on Evaluation, Standards, and Student Testing of The University of California at Los Angeles. (1990). Monitoring the impact of testing and evaluation innovations project: state activity and interest concerning performance-based assessment. Los Angeles: Author.
- Cizek, Gregory J. (1991). Innovation or enervation? performance assessment in perspective. Phi Delta Kappan, 72, 695-699.
- Darling-Hammond, Linda. (1991). The implications of testing policy for quality and equality. Phi Delta Kappan, 73, 220-225.
- Evans, Roxanne. (1992, June 1). Texas results involve culture, class more than race. Austin American-Statesman, p. 15.
- Feinberg, Lawrence. (1990). Multiple choice and its critics: are the "alternatives" any better? The College Board Review, 157, 13-31.
- Fink, Arlene, & Kosecoff, Jacqueline (1985). How to conduct surveys: a step-by-step guide. Beverly Hills: Sage.
-

- Frechtling, Joy A. (1991). Performance assessment: moonstruck or the real thing? Educational Measurement: Issues and Practice, 10, 23-25.
- Graves, Debbie. (1992, June 3). From classrooms to courtrooms: state mandated exams under fire. Austin American-Statesman, pp. 1, 4.
- Haney, Walter, & Madaus, George. (1989). Searching for alternatives to standardized tests: whys, whats, and whithers. Phi Delta Kappan, 70, 683-687.
- Kirk, Roger E. (1984). Elementary statistics (2nd ed.). Belmont, CA: Wadsworth.
- Linn, Robert L., Baker, Eva L., & Dunbar, Stephen P. (1991). Complex performance-based assessment: expectations and validation criteria. Educational Researcher, 20, 15-21.
- Mitchell, Ruth. (1992). Testing for learning: how new approaches to evaluation can improve American schools. New York: Free Press.
- Neill, D. Monty, & Medina, Noe J. (1989). Standardized testing: harmful to educational health. Phi Delta Kappan, 70, 688-697.
- Packer, Arnold H. (1992). Taking action on the SCANS report. Educational Leadership, 46, 27-31.
- Riverside Publishing Company. (1992). Norm-referenced assessment program for Texas coordinator manual. (Available from [the Riverside Publishing Company, 8420 Bryn Mawr Avenue, Chicago, IL, 60631]).

- Roeber, Edward. (1989, October). Performance assessment: it's real and feasible--but now it's critical. Paper presented at The University of California at Los Angeles Quality Indicators Conference, Los Angeles, CA.
- Selby, Gardner. (1992, April 5). It's multiple choice whether Texas' exams will stay. The Houston Post, p. 36.
- Shepard, Lorie A. (1989). Why we need better assessments. Educational Leadership, 46, 4-9.
- Stiggins, Richard J. (1987). NCME instructional module: design and development of performance assessments. Educational Measurement: Issues and Practice, 6, 33-42.
- Texas Education Agency. (1988, June). Rationale for the development of the student assessment program 1990-1995. Paper presented at the meeting of the State Board of Education, Austin, TX.
- Texas Education Agency. (1990, spring). Texas assessment of academic skills: mathematics objectives and measurement specifications 1990-1995. (Available from [the Texas Education Agency, 1701 N. Congress Avenue, Austin, TX, 78701-1494])
- Texas Education Agency. (1991, May). Texas assessment of academic skills and TEAMS exit level: student performance results October 1990. Austin, TX: Texas Education Agency.
- Texas Education Agency. (1992, April). Student assessment transition plan. (Available from [the Texas Education Agency, 1701 N. Congress Avenue, Austin, TX, 78701-1494])

Texas Education Agency. (1992, July). Press release of the associate commissioner for curriculum and assessment.

(Available from [the Texas Education Agency, 1701 N. Congress Avenue, Austin, TX, 78701-1494])

Texas Education Agency. (1992, summer). Texas assessment of academic skills: science objectives and measurement specifications 1990-1995. (Available from [the Texas Education Agency, 1701 N. Congress Avenue, Austin, TX 78701-1494])

Wiggins, Grant. (1989). A true test: toward more authentic and equitable assessment. Phi Delta Kappan, 70, 703-713.

Wiggins, Grant. (1991). A response to Cizek. Phi Delta Kappan, 72, 700-703.